

DRAFT

Evaluating the Reliability and Validity of The Health Plan Employee Data and Information Set

Luis Paita, Ph.D.
Denise Love, R.N., MBA
Kaveepan Lertwachara, MBA
Irina Grabovsky, Ph.D.

This study was funded by a grant made by the Society of Actuaries to the Office of Health Care Statistics (OHCS), Utah Department of Health. The authors acknowledge and thank the Society of Actuaries for its financial support for this research. The research and analysis is the work of the authors and does not reflect the views or policies of the Society of Actuaries.

Luis Paita, PhD and Denise Love, R.N., MBA, formerly with the OHCS, are now the Deputy Director and the Executive Director, respectively, of the National Association of Health Data Organizations (NAHDO). During the grant period, Kaveepan Lertwachara, MBA, University of Toronto, served as Research Consultant at OHCS and Irina Grabovsky, Ph.D., was Information Analyst at OHCS.

Direct all questions and correspondence to Luis Paita, PhD, Deputy Director, NAHDO, 391 Chipeta Way, Suite G, Salt Lake City, UT, 84108, lpaita@nahdo.org

DRAFT

EVALUATING THE RELIABILITY AND VALIDITY OF THE HEALTH PLAN EMPLOYEE DATA AND INFORMATION SET

TABLE OF CONTENTS

Executive Summary

1. Health Care Quality Assessment and Reporting	1
2. The Health Plan Employer Data and Information Set (HEDIS)	3
3. Research Method and Objectives	6
a. Research Objectives	6
b. Research Methodology	7
c. Research Data	9
4. Study Findings	13
a. Data Reliability Findings	13
b. Data Set Validity Findings	20
c. Content Validity Findings	23
d. Predictive Validity Findings	25
5. Conclusions	35

References

Appendices

Appendix A: Evaluation of Internal Consistency

Appendix B: Utah's HMO Satisfaction Survey: Methodology and Implementation

Appendix C: Glossary

DRAFT

EVALUATING THE RELIABILITY AND VALIDITY OF THE HEALTH PLAN EMPLOYEE DATA AND INFORMATION SET

EXECUTIVE SUMMARY

The Health Plan Employee Data and Information Set (HEDIS) is a set of standardized performance measures initially designed to enable purchasers of health care to evaluate and compare a managed care plan's administrative performance. The National Committee for Quality Assurance, known as the NCQA, developed HEDIS in 1991. Since its introduction, HEDIS has become a widely used tool for measuring a broad array of health plan functions.

This study evaluates the reliability and validity of HEDIS measures. It also reviews the usefulness of HEDIS as an indicator of health care quality.

▪ The Reliability of HEDIS Measures

The analysis of the reliability of HEDIS measures assesses the procedural reliability and the internal reliability of HEDIS measures. The procedural reliability assessment examines the consistency of HEDIS measures across different methods of measurements by comparing the HEDIS measures collected by three independent organizations for the same health plans. To assess internal reliability, the extent to which conceptually related HEDIS measures indicate similar levels of health plan performance, the correlation between specific measurements within comparable measure sets is examined.

▪ The Validity of HEDIS Measures

To test the validity of HEDIS measure sets, an analytic framework is used that postulates that there is a statistical relationship between health plan structural characteristics and enrollment

DRAFT

composition. Three broad categories of health plan performance, process, objective outcome and subjective outcome, are used to demonstrate whether there is such a relationship.¹ Process refers to health care services provided or medical interventions applied. Objective outcome refers to enrollees' health condition or health status. Subjective outcome refers to plan performance measured qualitatively.

To determine content validity of HEDIS measures which is the extent to which a set of HEDIS measures includes the universe of factors that should have been included in order to evaluate a health plan's performance, the relationship between a managed care plan's structural characteristics and the composition of its enrollment is compared with the plan's process measures. In addition, process measures and outcome measures are compared to determine HEDIS's predictive validity, the extent to which a set of HEDIS measures predicts an outcome or set of outcomes. Finally, process measures and subjective performance measures are compared to determine HEDIS's attributional validity, the extent to which the variation in enrollee satisfaction can be attributed to enrollment mix and structural characteristics as opposed to health plan performance.

▪ **Reliability and Validity Findings of HEDIS Measures**

Strong agreement on HEDIS measures for the same set of health plans indicated high procedural reliability.

High internal reliability was observed within individual plan's HEDIS reports, such that health plans that perform well according to one measure tend to perform well based on other measures in the same set. However, the consistency between measure sets was weaker. Overall, a weak correlation was observed between process measures and corresponding outcome measures. Nevertheless, the study indicated that subjective measures are a useful indicator of a health plan's performance. The study identified areas of high internal consistency and

¹ These categories may also be referred to as "domains."

DRAFT

opportunities for simplification of HEDIS through strategic reduction of measures and calculation of composite scores. Additionally, the study suggests the need for encounter level data to supplement HEDIS analysis.

- **HEDIS' Utility in Quality Assessment**

Finally, the analysis supports the use of HEDIS as a tool for quality of care improvement. The availability of HEDIS information enables development of national, long-term, comparable, databases. The availability of such data allows for analysis of the many different factors that contribute to quality of care and the effective provision of high quality health care services.

DRAFT

EVALUATING THE RELIABILITY AND VALIDITY OF THE HEALTH PLAN EMPLOYEE DATA AND INFORMATION SET

1. HEALTH CARE QUALITY ASSESSMENT AND REPORTING

Health care quality assessment has been a subject of discussion and debate. Policy analysts have struggled to find consistent, universally accepted ways to define and measure quality. In recent years, as health care delivery and its financing have become intertwined, quality assessment has become increasingly important.

Under managed care, quality of care—once solely the responsibility of physicians and hospitals—has become the responsibility of a multitude of health care system participants. When a health plan “manages” care, quality assessment and quality assurance efforts must address all aspects of administrative and clinical performance. This ranges from third-party payers’ claims processing procedures and provider panel selection to the enrollees’ health outcomes and satisfaction with services and care received. With so many health care system participants influencing health care delivery and its financing, it has become increasingly difficult to hold any one participant accountable for the health plan’s quality of care and administrative performance. Healthcare purchasers, whether individual consumers, large corporate benefits managers or government agents, need indicators of quality to make prudent choices among health plans. Thus, there is need for and interest in reliable, understandable, information about all aspects of health care quality.

Managed care plan performance information can assist the efforts to establish greater accountability for the quality of care and the services provided through managed care arrangements as well as for a managed care plan’s administrative performance. Indeed, many analysts believe that the proliferation of performance measurement initiatives² will help mandate

² Faulkner and Gray, 1997 Comparative Performance Data Sourcebook; 1998 Comparative Performance Data Sourcebook; 1999 Comparative Performance Data Sourcebook. New York, NY: F & G Inc. 1997, 1998 and 1999, respectively.

DRAFT

universal accountability. With accurate performance information, it is thought that health care providers can be held accountable for the clinical quality of the care they provide and managed care plans could be held accountable for their enrollees' access to care. Furthermore, employers can be held accountable for the health plans offered to their employees and individual consumers can be held accountable for their health plan choices and for their behaviors that impact health risk.

DRAFT

2. THE HEALTH PLAN EMPLOYER DATA AND INFORMATION SET (HEDIS)

During the past 10 years, major health plan purchasers and national managed care plans have embraced the Health Plan Employer Data and Information Set (HEDIS) as a tool for plan performance measurement and reporting. The data set, referred to as HEDIS, was developed by the National Committee for Quality Assurance (NCQA), a not-for-profit organization committed to evaluating and publicly reporting on the quality of managed care plans. The NCQA introduced HEDIS in 1991.

HEDIS helped health plans begin to meet the emerging market demand for standardized performance reporting. Major employers, stung by the health care benefit cost increases of the 1970s and 1980s, hoped to gain tighter control of health care benefit expenditures by holding health care suppliers to production and cost standards—treating health care as they would any other cost of production. Large national employers, many of which were manufacturers, wanted their health plans to provide them with the same kind of administrative performance measures that they were accustomed to receiving from their other large suppliers.

Since introducing the first iteration of HEDIS, HEDIS 2.0 in 1993, the NCQA has modified and expanded the data set to reflect the growing sophistication and data management capabilities of health plans and to respond to purchasers' increasing interest in outcomes and cost management information. In 1996 the NCQA introduced HEDIS 2.5, which included supplemental measures specifically targeted to Medicaid and Medicare. The NCQA introduced HEDIS 3.0 in 1997.³

HEDIS 3.0 combined earlier versions of HEDIS into one integrated, comprehensive data set to accommodate Medicaid, Medicare, and commercial populations. HEDIS 3.0 reduced the

³ HEDIS 3.0 had been released at the time of this study, however available data was based on HEDIS 2.5.

DRAFT

reporting burden that health plans faced by reducing multiple reporting specifications into a single set and by increasing the population size to be evaluated. In 1999 the NCQA offered a further revised edition, HEDIS 2000.

By 1994, about 75% of health plans assessed their performance using one or more HEDIS measure(s).⁴ By 1997, 60% of HMOs had implemented and begun to publish HEDIS based reports.⁵ HEDIS information now is widely used to report health plan performance regarding administrative management and clinical record keeping.

▪ **HEDIS Data Collection: A Challenge for Managed Care Organizations**

The resources and infrastructure required to collect and report accurate HEDIS measures are extensive. Many managed care organizations lack the systems and data collection, management, and analytic infrastructure needed to support HEDIS. A plan's capability for data collection to construct accurate and comparable HEDIS measures may be further challenged by private market merger and acquisition activity that alter enrollment numbers and administrative policies and procedures.

A 1997 managed care systems capabilities assessment conducted by the NCQA, of Utah's leading HMOs for the Utah Department of Public Health, Office of Health Data Analysis, identified the following problems:

- Varying degrees of health plan experience with HEDIS reporting;
- Balancing information system upgrades with consistency in data systems;
- Mergers and buyouts resulting in potential systems incompatibility and the exclusion of newly enrolled populations;
- Vendor contracts without adequate data quality assurance procedures in place;

⁴ A. Epstein; "Rolling Down the Runway. The Challenges Ahead for Quality Report Cards." In *JAMA*, June 3, 1998, 279(21), 691-1696.

⁵ National Committee for Quality Assurance, HEDIS 3.0, vol.1, NCQA, Washington DC, 1997.

DRAFT

- Carve outs and global rates resulting in incomplete or missing data;
- Internally developed coding schemes and variable use of automated coding software;
- Unedited encounter data in some utilization databases;
- Various product lines combined into one HEDIS report;
- Miscounts of denominator populations due to enrollment tracking problems; and,
- Data entry errors.

As HEDIS measures become a widely used reporting tool, a systematic evaluation of the extent to which the measures provide a reliable basis for health plan comparison is needed.

DRAFT

3. RESEARCH OBJECTIVES AND METHODOLOGY

▪ **Research Objectives**

This study evaluates, first, whether HEDIS measure sets provide reliable and valid indicators of performance. The research evaluated HEDIS 2.5 measures against five important characteristics:

- Procedural Reliability: The extent to which HEDIS measures are consistent across different methods of measurement.
- Internal Reliability: The extent to which conceptually related HEDIS measures indicate similar levels of health plan performance.
- Content Validity: The extent to which a set of HEDIS measures includes the universe of factors that should have been included in order to evaluate a health plan's performance.
- Predictive Validity: The extent to which a set of HEDIS measures predicts or correlates with an outcome or a set of outcomes.
- Attributional Validity: The extent to which the variation in enrollee satisfaction with the health plans can be attributed to enrollment mix and structural characteristics, as opposed to health plan performance.

Second, the analysis seeks ways that HEDIS data collection might be made less burdensome—yet equally instructive. The analysis asks:

- Is annual compilation of all HEDIS measures necessary?
- Would it be equally instructive if some measures were collected more often, and some less frequently?
- Should rotation of select measures be considered and, if so, which measures and what would be the optimal mix and timing of measures?
- Could other data serve as proxies for HEDIS measures?

DRAFT

Finally, the research looks at the extent to which enrollee satisfaction survey results reflect health plan performance, independent of enrollee characteristics.

- **Research Methodology**

- **Evaluation of Reliability**

The reliability of HEDIS measures refers to the consistency, stability, or equivalence of its results: 1) over time; 2) between repeated or related measures of the same concept (internal reliability); or, 3) across different methods of measurement, data sources and entities conducting the measurement, (procedural reliability).⁶ This study looks only at internal and procedural reliability.

The analysis evaluated procedural reliability by examining the consistency between the HEDIS measures collected by three independent organizations. The research examined the internal reliability of HEDIS measures using Cronbach's alpha. Cronbach's alpha ranges from 0 to 1.0 and indicates how much individual measures in a set are measuring the same characteristics.⁷ The results of the internal reliability analysis provided the basis for consolidating measures into groups of related measures, or an overall index of performance for the evaluation of validity.

- **Evaluation of Validity**

Validity is a multidimensional concept. In this analysis validity is the degree to which there are systematic differences between a set of measures relative to the full meaning of the concept it was intended to express; i.e., health plan performance, related measures of performance, or theories or hypotheses about their relationships to other concepts .⁸

⁶ At the time of this study, HEDIS was relatively new and the data set needed to perform sufficient analysis of its reliability over time was not accessible to the investigators.

⁷ Vogt, 1993, p 53; Carmines and Zeller, 1979, p. 44-51.

⁸ Aday, 1989, p.47.

DRAFT

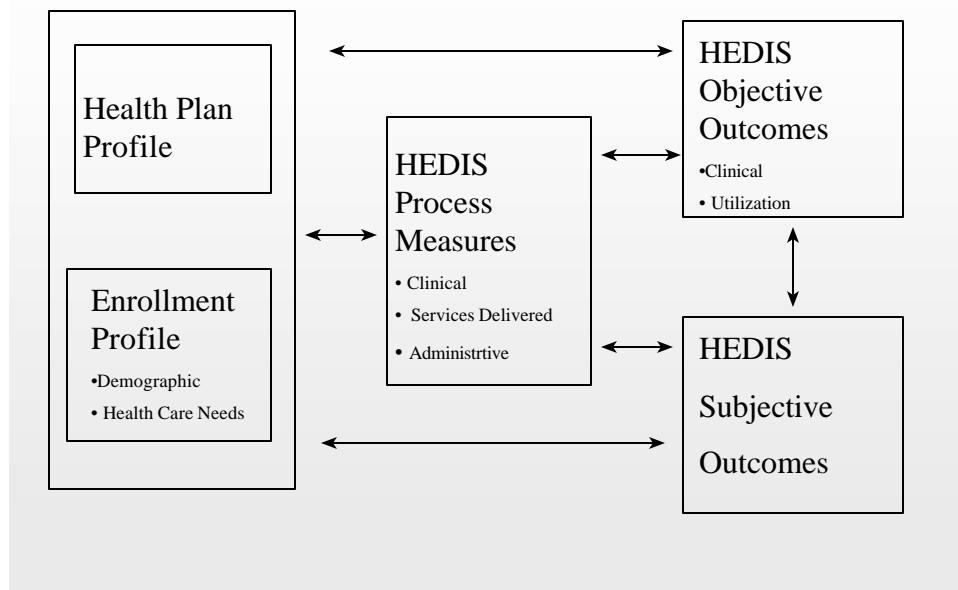
The analytical framework is depicted in Figure 1. The validity phase of the research followed this framework.

To place the analyses within a conceptual framework, a distinction is made among the following characteristics and types of measures:

- Health Plan Profile - characteristics of the health plan in terms of its structure (e.g., financial condition, provider profile);
- Enrollment Profile - characteristics of its enrollees (e.g., age, sex) which serve as the exogenous factors in the model, and are hypothesized to be factors affecting the process and outcome measures.
- Process Measures - Measures that refer to the volume of services delivered by the health plan (e.g., percent mammography); and
- Outcome Measures - Measures that refer to either objective (e.g., incidence of adverse conditions among enrollees), or subjective (i.e., enrollee satisfaction) outcomes.

DRAFT

Figure 1 - Study Framework



C. Research Data

The analysis used three independent sources of HEDIS 2.5 data: Health Benefits America (HBA), William M. Mercer and Co., and the NCQA. Only the HBA dataset was used in the assessment of internal reliability and validity. The other two sources were used only for comparison with HBA as proxy for assessing procedural reliability. In addition, the analysis used the State of Utah's 1996 HMO Enrollees Satisfaction Survey to analyze enrollee satisfaction.

o Health Benefits America Data

A Health Benefits America (HBA) data set provided health plan descriptive information, complete HEDIS 2.5 measures, satisfaction survey results, and other measures. The completeness of the HBA data set allowed data transformation and further derivations of

DRAFT

measures to compare and contrast health plan performance across different organizational structures and locations. Table 1 illustrates the distribution of HBA data across geographic areas (Midwest, West, Northeast and Southern regions) and managed care models -- Staff Model, Independent Practice Association (IPA) Model, Network Model, Group Practice Model and Mixed Models.

- **William M. Mercer, Inc.**

The William M. Mercer, Inc. data set contained 27 HEDIS measures and was examined for consistency with the other two sources. Most Mercer HEDIS measures were process measures with some outcomes, structural indicators, and overall enrollee satisfaction results. Of the 27 measures, only one—asthma admission rates—could not be directly compared to HBA and NCQA databases because the age groups for the measures were not comparable. The percent of health plans that reported HEDIS measures in the Mercer data set ranged from 7% to 32%. However, the total number of health plans that reported each of the 27 HEDIS measures in the Mercer dataset ranged from 60 plans (the second lowest number was 135 plans) to 266 plans. Among the 838 health plans in the Mercer database, 255 reported at least 10 HEDIS measures and 331 plans reported at least one HEDIS measure.

- **NCQA Quality Compass**

Quality Compass is a national database created by the NCQA based on the HEDIS 2.5 measures and accreditation information of 250 health plans. Quality Compass data enables direct, plan-to-plan comparisons on a variety of different measures ranging from effectiveness of care to selected clinical procedures. However, Quality Compass does not provide detailed enrollment information across age groups, genders and payers, needed for a health plan's structures in this analysis. Therefore, Quality Compass measures were used for the examination of consistency with the HBA data only.

DRAFT

Table 1
Health Benefits America
Data Distribution (Valid N=253)

Regions	Percentage
Northeast	23%
Midwest	26%
South	24%
West	28%

Model Types	Percentage
Mixed	31%
IPA	48%
Others (Network, Staff, Group)	15%
Not Reported	6%

Table 2
NCQA Quality Compass Data Distribution (Valid N=211)

Regions	Percentage
Northeast	24%
Midwest	24%
South	31%
West	21%

Model Types	Percentage
Mixed	33%
IPA	49%

DRAFT

Others (Network, Staff, and Group)

18%

- **1996 Utah Survey of Enrollees in Medicaid-Contracted HMOs (Enrollee Satisfaction Survey Data)**

The 1996 Utah Survey of Enrollees in Medicaid-Contracted HMOs provided the enrollee satisfaction data for the analysis.⁹ The survey relied on telephone interviews with 2,217 Medicaid enrollees for the five HMOs with Utah Medicaid contracts.

Survey respondents were the named beneficiaries in the target population aged 18 years or older, those with experience with their health plan and those who had been Medicaid enrollees for at least three months prior to the survey.

The study used six variables to reflect social location or position: age, gender, education, income (socio-economic status), race, and ethnicity. Several indicators were used to measure health status including general perception of health, self-reported physical health, self-reported mental health, serious chronic conditions, and other chronic conditions. Respondents' health care experience was derived from health care utilization and from responses to questions about problems gaining access to care.

The survey asked respondents to rate their health plan or care in 22 specific areas. Using factor analysis, the 22 items were reduced into five domains of care: quality of care, coverage, appointments, provider choice, and physical access. Subsequent analyses, performed after the domains had been created, used composite scores that had been calculated for these five domains.

⁹ Conducted by the Utah Department of Health

DRAFT

4. STUDY FINDINGS

1. Reliability Findings

A. Consistency Among Data Sources (Procedural Reliability)

- **Comparison between Quality Compass and HBA HEDIS Data**

Comparison of the distribution of data across geographic regions and model types showed that the NCQA Quality Compass and the HBA HEDIS data sets contained similar information. Samples of the plan-level comparison provided in Appendix A indicate the level of consistency of HEDIS reporting between these two data sets.

Of 100 health plans in the HBA database, 34 also reported HEDIS measures in the NCQA Quality Compass database. More than 90% of the data reported in both databases had the same values. A sample of performance measures of Utah, Arizona, and California health plans is presented in Table 3.

While the Quality Compass database provides limited information on plan enrollment and structure, the HBA database contains enrollment information by age group, payer, and gender. Moreover, with the HBA database, analyses can be performed adjusting for age and gender for almost every performance variable.

DRAFT

Table 3
Samples of Plan-Level Comparison

Health Plan A of Utah	HBA	QC
Diabetic Retinal Exam	48.1%	48.2%
Cervical Cancer Screening	71.0%	68.6%
Cholesterol Screening	74.1%	74.1%
Mammography	89.3%	69.5%
Prenatal Care in the First Trimester	91.0%	91.5%
Health Plan B of Arizona	HBA	QC
Cholesterol Screening	78.8%	78.8%
Mammography	75.8%	75.8%
Cervical Cancer Screening	86.4%	86.4%
Prenatal Care in the First Trimester	87.5%	87.5%
Asthma Inpatient Admission Rate 2-19 yr old	00.7%	00.7%
Asthma Inpatient Readmission Rate 2-19	00.1%	00.1%
Asthma Inpatient Admission Rate 20-39	00.1%	00.1%
Asthma Inpatient Readmission Rate 20-39	00.2%	00.00%
Diabetic Retinal Exam Rate	38.8%	38.8%
Ambulatory Follow-up After Hospitalization for Major affective disorder	75.5%	NR
Health Plan C of California	HBA	QC
Cholesterol Screening	79.1%	79.1%
Mammography	65.9%	65.9%
Cervical Cancer Screening	75.1%	75.1%
Prenatal Care in the First Trimester	90.4%	90.4%
Asthma Inpatient Admission Rate 2-19 yr old	00.8%	0.8%
Asthma Inpatient Readmission Rate 2-19	00.2%	0.2%
Asthma Inpatient Admission Rate 20-39	00.2%	0.2%
Asthma Inpatient Readmission Rate 20-39	00.0%	0.04%
Diabetic Retinal Exam Rate	35.9%	35.9%
Ambulatory Follow-up After Hospitalization for Major affective disorder	23.3%	23.3%

A sample of administrative process measures, including actual waiting time for urgent care visits, emergency visits, urgent Mental Health/Chemical Dependency (MH/CD) care visits, emergency MH/CD care visits, phone access to customer service, and phone access to MH/CD customer service, shows a significant difference in the completeness of the NCQA and HBA data. While HBA reports 38% missing values in this category, Quality Compass contains more

DRAFT

than 57% missing values. Table 4 indicates a higher percentage of usable data in the HBA data set.

Table 4
Percent of Missing Values

Performance Measures	HBA	Quality Compass
Administrative Process*	37.75%	57.35%
Utilization – Maternity & Mental Health	2.94%	15.44%**
Clinical Outcomes	12.18%	14.12%***

* Administrative process measures include actual waiting time for urgent care and emergency care and actual waiting time for phone access to customer services.
** No length of stay for mental health services reported in Quality Compass.
*** Limited readmission rate to within 365 days. No readmission within 90 days reported.

○ **Comparison between William M. Mercer, Inc. and Health Benefits America HEDIS Data**

Of 253 plans in the HBA database, the HEDIS data of 101 plans was also included in the Mercer database. Because the Mercer data set includes only 27 HEDIS measures, it was used only for validation purposes. Overall, 56% (Ranges from 34% to 74%) of the HEDIS measures reported by health plans in the Mercer data set match the HEDIS measures reported by those same health plans in the HBA database. Table 5 lists HEDIS measures provided by Mercer and the percent of matching values.

DRAFT

Table 5
HEDIS Measures

Measure	Percent Matched Between Mercer and HBA Databases
DPT	59.46%
OPV	62.16%
Measles	53.15%
Mumps	51.35%
Rubella	49.55%
Influenza b	62.16%
Overall Childhood Immunization	63.06%
Cholesterol	58.56%
Mammography	69.37%
Cervical Cancer	62.16%
Prenatal Care	64.86%
Diabetic Retinal Exam	64.86%
Very Low Birth Weight	73.87%
Low Birth Weight	67.57%
Board Certified PCP	56.76%
Board Certified Specialists	62.16%
Provider Turnover	50.45%
Overall Satisfaction	63.06%
Mental Health Follow up	34.23%

DRAFT

Overall, HBA data contains information similar to the other two independent data sources. Because HBA data provides additional information, it is the only data set that could accommodate this analysis of the reliability of health plan performance measures.

B. Consistency Within Measure Sets (Internal Reliability)

A measure set is defined as a group of measures that pertain to a particular domain or characteristic of a health plan or health care received. In most cases, measure sets correspond to the NCQA categorization of measures. For example, effectiveness of care comprises a measure set.

Evaluating internal reliability involved examining whether the relative ranking of health plans remains consistent across different HEDIS measures within measure sets. In other words, do health plans that score high in one measure also score high in other measures in its set? Internal consistency refers to strength of correlation among measures within measure sets. The measure sets examined for this study are listed in Appendix A.

Internal consistency was examined in three ways. First, the correlation matrix containing pair wise correlations between measures in the set was examined for each measure set. Instead of the absolute values of the measures, non-parametric correlation (Spearman's Rho) was used because of the non-normal nature of most variables and interest in looking at relative ranking.

Second, correlations based on the original units of measures were calculated for homogeneous sets of measures, that is, sets of measures that are based on the same unit of measure; for example, rates per 1,000. For nonhomogeneous sets in which different units of measures are used, calculations were based on relative rank transformed values. For example, for a measure set called resource utilization intensity, which contained a combination of average cost, length of stay and discharge rate, the values were first transformed into relative ranks prior to calculations. Next, Cronbach's alpha reliability coefficient was calculated.

DRAFT

The value of Cronbach's alpha indicates whether the measure sets had high, moderate or low internal consistency. The higher the value of alpha, the greater the internal consistency. Thus, 0.8 and above indicated a high internal consistency, between 0.6 and 0.8 indicated moderate consistency, and less than 0.6 reflected low internal consistency. Finally, factor analysis was used to validate the results of the correlation matrix analysis and the alpha and to identify smaller sets of measures within that set that show high internal consistency.

The results for the measure set of immunization rates, which is comprised of seven measures (overall immunization and six type-specific immunization rates) illustrate this approach. All pair-wise Spearman correlation coefficients calculated for this measure set were significantly high with p less than 0.001. The number of valid values (the number of health plans with values for each pair of measures) ranged from 150 to 199. The correlation matrix analysis indicates a high internal consistency within this measure set. The calculated alpha value of 0.95 leads to the same conclusion. Factor analysis on the entire effectiveness of care set shows type-specific immunization rates loading together as one factor (Researcher: What does loading together mean?), indicating again a high internal consistency in that set of immunization rates.

This process was repeated for several measure sets and categorized according to high, moderate and low levels of internal consistency.

Most of the measure sets were highly internally consistent. Included in these measure sets were three subjective measures: satisfaction rate, percent willing to recommend, and retention rate (complement of disenrollment rate). Also in this high internal consistency group are immunization rates (as previously described), high-cost diagnosis-related groups (DRGs) selected procedures, all the length of stay measures, and mental health-related utilization measures and utilization rates for chemical dependency (MH/CD). All the mental health related measure sets are highly internally consistent not only within each set but also between sets. That is, mental health utilization measures were highly correlated with chemical dependency

DRAFT

utilization, readmission for major affective disorder and mental health readmission rates. In fact, the measures in all sets together comprise a highly internally consistent set.

In addition to the measure sets listed above, subsets of measures examined showed high internal consistency. Within effectiveness of care two subsets were asthma admission and readmission and preventive screening rates (mammography, cholesterol screening and cervical cancer screening).

The study found that measures reported for specific age groups of enrollees were highly internally consistent. HEDIS inpatient, outpatient, emergency room, and ambulatory surgery utilization rates, for example, are calculated for age groups 0 to 19, 20 to 44, 45 to 64, and 65+. All of these sets were found to be highly internally consistent, meaning that health plans with high rates for one age group generally tend to have high rates for the other age groups. In addition, the all-ages measure had the strongest and most consistent correlation with the age-specific measures. Therefore, when comparing utilization rates between health plans, it would be sufficient to use the overall measure without looking at the age breakdowns, significantly simplifying the analysis. This is also true of maternity and newborn-related discharges, which are calculated for three age groups: 10-14; 15-34; and 35-49.

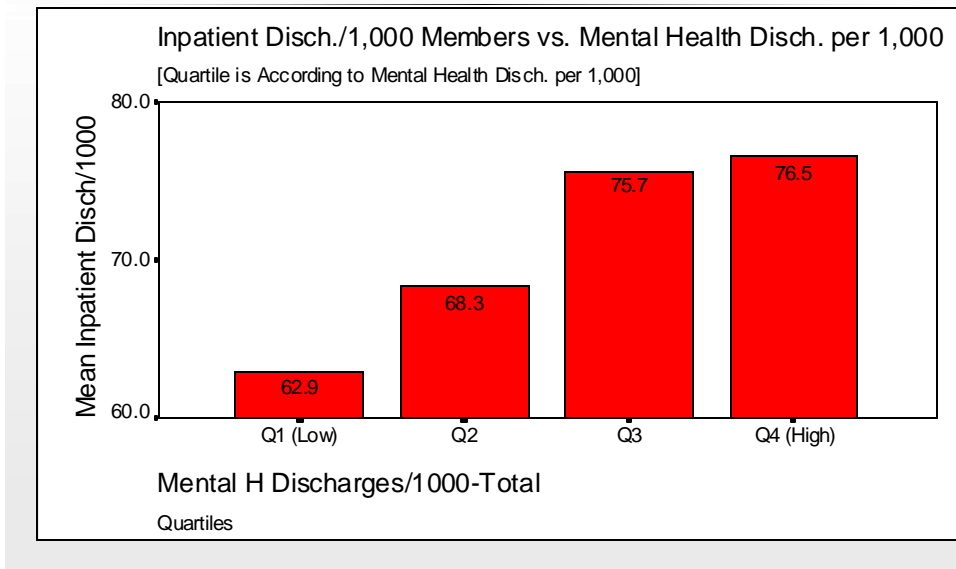
It also was found that for measures reported separately for males and females the overall measure represents each set adequately. This is true for utilization rates for selected DRGs, utilization for mental health conditions, readmission for major affective disorder, chemical dependency utilization, and readmission for chemical dependency.

Figure 2 illustrates a highly consistent pair of measures. This chart shows the relationship between rate of inpatient discharges per 1,000 people and rate of mental health discharges per 1,000 people. The x-axis reflects the quartiles for mental health discharge and the y-axis shows the average for the inpatient discharge rates. It shows that health plans in the higher quartiles of mental health discharge rates are also the ones with the high inpatient

DRAFT

discharges per thousand. This monotonic, positive correlation is representative of most of the pair-wise correlations looked at for sets that are highly internally consistent.

Figure 2 –Example of High Internal Consistency



Although most of the measure sets were highly consistent, some showed moderate internal consistency (Alpha between 0.6 and 0.8). This includes the preventive screening set (cholesterol, mammography, cervical cancer), asthma admission and readmission rates, high cost/occurrence DRGs, total deliveries, vaginal deliveries, and maternity care length of stay.

The results of the evaluation are provided by measure set in Appendix B.

DRAFT

2. Data Set Validity Findings

Validity was assessed by examining relationships between exogenous measures, process measures, objective outcome measures, and subjective outcome measures (again, see Figure 1). Each domain is described briefly below.

A. Exogenous Measures

o Health Plan Characteristics

“Health plan characteristics” refers to an aggregate description of health plans. Structural measures included in this study were HEDIS descriptive information (provider profiles, health plan management, or financial position of health plans) and equivalent information of other health plan settings (e.g., indemnity). This information and data set is necessary for the cross-structural analysis, in which performance measures are tested across various types of health plan settings. Cross structural analysis could provide critical adjustments to improve comparability.

Four structural measurements are used in the study, primarily as confounding factors.

- 1) Organizational stability
- 2) Provider profiles
- 3) Model types
- 4) Geographic areas of service

The study examined indicators of organizational stability:

- 1) Change in operating profit margin;
- 2) Change in net income;
- 3) Change in member months by product line;

DRAFT

- 4) Change in member months by year in business;
- 5) Change in member months by licensure;
- 6) Change in member months by loss ratio; and
- 7) Change in member months by change in loss ratio.

Provider profile measures included:

- 1) Percent of primary care physicians (PCPs) with a completely open panel;
- 2) Percent of PCPs with a completely closed panel;
- 3) Percent of PCPs with restrictions on additional members;
- 4) Number of PCPs with open panels per 1,000 members;
- 5) Number of PCPs with closed panels per 1,000 members;
- 6) Number of PCPs with restrictions on additional members per 1,000 members;
- 7) Percent of PCPs who are board certified;
- 8) Percent of specialists who are board certified; and,
- 9) PCP turnover rate.

Each HMO included in the study is classified as a staff, IPA, network, group, or mixed model and each HMO also is identified with its region--Midwest, West, Northeast, or South.

o **Enrollment Composition**

“Enrollment composition” describes global information about the enrollment mix in health plans. HEDIS measures provide some demographic information about enrollees, including age, sex, and duration of enrollment. Additional information regarding general health status of enrollees, chronic health conditions, or ethnicity collected from other sources would add to risk adjustment analysis. The percentage of total enrollment in each of the following four categories was used to describe composition:

- 1) Age, from newborn to 85, in four-year groups;
- 2) Payment source—direct pay/group plan;

DRAFT

- 3) Medicaid and Medicare by age group; and
- 4) Gender.

B. Process Measures

The key objective of the research was to describe how these process measures are related to other performance indicators. The study in this particular area consists of three levels: individual measure, domains of measures, and health plans' overall performance. At the individual level, analysis was conducted to evaluate predictive values of performance measures over one another. For each domain of health plan performance, the study tried to find major contributing measures and evaluate the possibility of proxy measures that could represent the entire domain of care.

a. Objective Outcome Measures

Objective outcome measures include outcome-related performance measures, such as length of stays and clinical outcome rates. These measures are analyzed to explain the extent of the relationship between clinical outcomes and a health plan's organizational or structural characteristics. In addition, predictive value of process measures over clinical outcomes was assessed. For example, cesarean-section rate and average length of stays for maternity and newborn services were studied in relation to relevant process measures, such as initiation of prenatal care, prenatal care during the first trimester, and frequency of ongoing prenatal care.

b. Subjective Outcome Measures (enrollee satisfaction)

The NCQA's HEDIS measures include the enrollee's satisfaction survey. The Utah Office of Health Care Statistics conducted the 1996 and 1997 Utah survey of consumer satisfaction with their HMOs. Survey findings confirm the relationship between consumer satisfaction and performance indicators.

DRAFT

Subjective outcome measures are further studied as a separate module. The data utilized in this module were obtained from Utah's 1996 Survey of HMO Enrollees and HBA/Mercer HEDIS satisfaction survey results.

3. Content Validity Findings

Content validity was assessed by examining the correlation between health plan characteristics and enrollment profile, and by examining the correlation between HEDIS process and outcome measures. A significant correlation may indicate weak content validity for purposes of comparing health plan quality because variations in HEDIS measures would reflect variations in health plan characteristics and enrollment composition--rather than performance. The results of this assessment indicate whether comparing health plan performance according to HEDIS measures would require adjustment, and if so, for which health plan or enrollee characteristics.

The health plan characteristics included indicators of structure and stability. After preliminary analyses, which included a series of internal consistency checks, the following were included: health plan type, regional location, percent of primary care physician with open panel, percent change in total profit margin, percent female, percent aged 0-19, and percent aged 65+. The matrix in Figure 3 summarizes the results of the analysis of variance, comparing the means of HEDIS measures (rows) between groups of health plans defined by health plan and enrollee characteristics (columns). Cells marked with X indicate statistical significance of the calculated F-ratio, indicating significant difference in the means for the measure (row) between at least two groups defined by the variable on the column. The matrix shows some evidence of the need to take into consideration health plan characteristics and enrollment profile, especially for age and sex composition. This could be achieved through direct or indirect adjustment methods, or through stratification of the comparative analysis.

DRAFT

Figure 3 – Health Plan Characteristics and Enrollment Composition vs HEDIS Measures

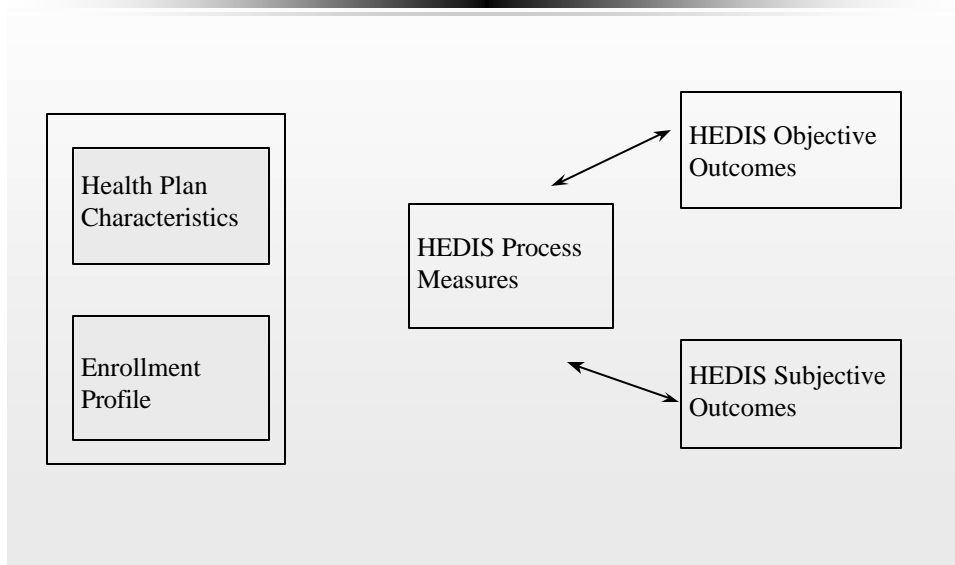
Measures	Plan type	Plan's Region	Percent PCPs w/ Open Panels	Percent of Female Enrollees	Percent of enrollees 65+	Percent of enrollees 0-19
- Overall Childhood Immunization	X	X	X	X		X
- Composite Asthma Measure			X	X	X	
- Average Preventive Screening Rate	X			X	X	
- Heart-related DRGs:discharges/1000				X		
- I/P discharges/1000			X			X
- O/P Visit /1000		X		X	X	
- Amb.surg. proc/1000			X		X	X
- ER Visits/1000	X	X		X	X	
- Mental Health discharges / 1000		X		X	X	X

4. Predictive Validity Findings

Predictive validity was examined by looking at the relationship between selected HEDIS process measures and corresponding outcome measures. High predictive validity means that health plans that perform well in process measures are also expected to perform well in outcome measures. Figure 4 illustrates the relationships examined.

DRAFT

Figure 4 - To what degree do HEDIS process measures correlate with performance in outcome measures?



First, the degree to which a health plan's relative performance in HEDIS process measures serves as predictor of its relative performance in objective outcome measures was reviewed. The findings suggest a strong correlation between immunization rates and the following rates: pneumonia and pleurisy; age group 0-19 bronchitis and asthma discharges, average cost and length of stay; inpatient non-acute discharges and length of stay for age group 0-19.

A strong relationship was found between preventive screening measures (mammography, cholesterol, cervical cancer) and the following utilization measures: inpatient length of stay, outpatient visits per thousand, emergency room visits per thousand, and inpatient non-acute days per thousand.

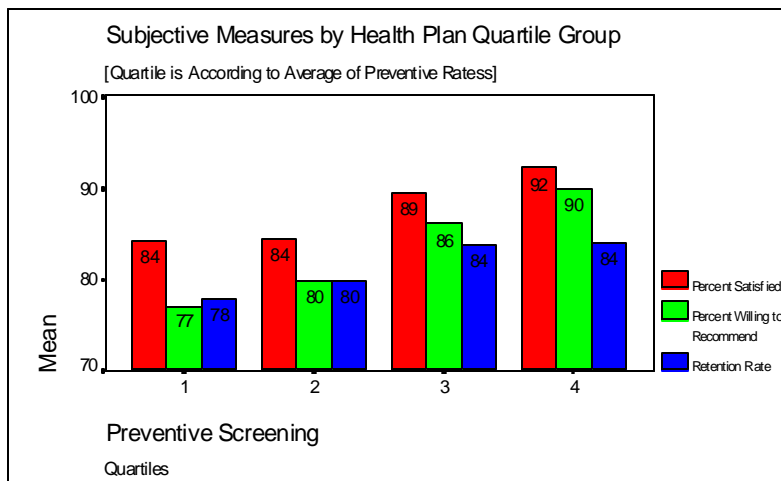
Also interesting is the absence of statistical relationships between particular pairs of process and outcome measures that were hypothesized to have strong correlations. The most notable are: ambulatory follow-up for major affective disorder (MAD) and readmission for

DRAFT

MAD; cholesterol screening and heart-related DRG utilization; cervical cancer screening and hysterectomy; and prenatal care and low birth weight.

A significant correlation was found between some HEDIS process measures and subjective performance measures, as illustrated in Figure 5, for preventive screening. Subjective outcomes are satisfaction, recommendation, and retention rates. Satisfaction rate was influenced by overall immunization rate, composite preventive screening (average of mammography, cholesterol screening, and cervical cancer screening rates), and the following health care utilization measures: inpatient stays for ischemic attack and nonspecific cerebrovascular diseases, laminectomy, outpatient visits, ambulatory surgery, and maternity days. Strong relationships were observed between retention rate and overall immunization rate, composite preventive screening, length of stay for nonspecific cerebrovascular diseases, outpatient utilization, newborn inpatient average length of stay (ALOS) and maternity days.

Figure 5 – Example of High Predictive Validity , Process Measures vs Subjective Outcome Measures



In general, low predictive validity was found for the measures examined. These results may reflect the fact that the unit of analysis is health plan rather than individual enrollee. Causal

DRAFT

relationships that might be expected on an individual level may not hold true at the health plan level. Rates expected to be strongly related may show weak correlation empirically due to the fact that the rates are calculated based on the experience of different subgroups of enrollees within a health plan. Additionally, the measures were calculated for the same time period whereas the hypothesized relationships may be based on a time lag (e.g., effect of prenatal care on birth weight).

5. Attributional Validity Findings

Some significant variation in HEDIS measures between health plan groups defined by health plan characteristics and by enrollee composition was noted earlier. This suggests the need to adjust for these characteristics in the model.

Multiple regression of satisfaction rates and retention rates found the HEDIS measures to be strongly correlated with the measures described above, with selected health plan and with enrollee characteristics in the model. The satisfaction rate R-square of 0.68 shows significant effects of immunization rate, outpatient visits, ambulatory surgery, and maternity days. The retention rate R-square of 0.31 indicates that composite preventive screening, maternity days, and newborn inpatient ALOS were statistically significant.

These results suggest that variation between health plans' subjective outcomes reflects variation in their performance in HEDIS process and objective outcome measures, beyond those that can be attributed to the type of health plan or the composition of their enrollees. This finding alone is interesting, but converting it into practical use requires understanding the nature of the implied causal relationships.

Unfortunately, there are limitations when examining causation between aggregate measures like HEDIS. First, the target population varies from measure to measure. For example, satisfaction rates may be derived from a sample of enrollee responses representative of

DRAFT

a subgroup of the health plan's enrollee population that is different from that where immunization rates are derived.

Second, the effect of a health plan's performance in a "process" aspect of care may affect its performance in a related "outcome", perhaps only after some period of time has lapsed. This is illustrated by the absence of a statistically significant relationship between prenatal care rate and low birth weight rate.

Enrollee-level data could provide more useful information regarding the factors that affect a health plan performance when measured at the aggregate level. A parallel analysis using enrollee level data from Utah's survey of enrollee satisfaction with HMOs was performed using a framework similar to that described earlier, using satisfaction measures as the dependent variables. Satisfaction was defined to include overall satisfaction and satisfaction with five domains of health plan or care: quality of care, coverage, appointments, provider choice, and physical access. These domains were derived through factor analysis of 22 satisfaction survey items.

Included as factors that drive satisfaction are enrollee characteristics or social location, health status or health care needs, and enrollee health care experience. Of primary interest were enrollees' health care experiences, measured in terms of both health care utilization and perceived problem with access (a composite of problem with delay in care, referral, and obtaining needed care). This enabled an examination of whether enrollee satisfaction was:

- 1) Influenced by their actual experience with the health plan;
- 2) Beyond the influence of their socio-demographic position; or
- 3) Influenced by health status.

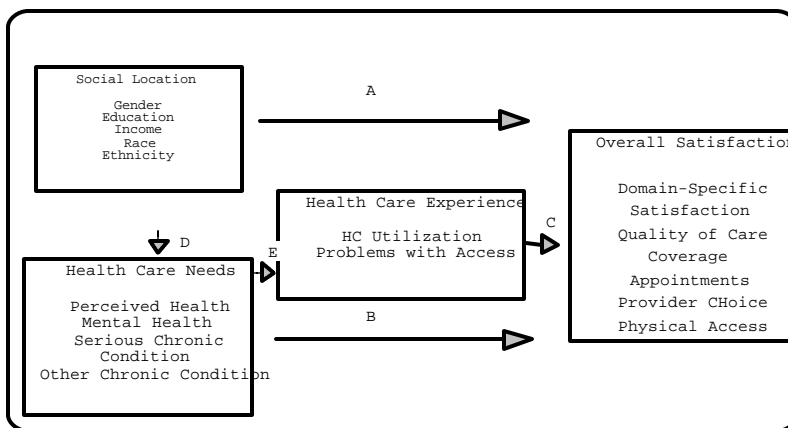
The analysis was based on data collected by the Office of Health Data Analysis through the 1996 Survey of Enrollees in Medicaid-Contracted HMOs.¹⁰ (See Appendix C for a detailed discussion of the methodology and survey implementation.)

¹⁰ Paita, 1997

DRAFT

Figure 6 shows the hypothesized relationships examined in the study. Of interest were interested in examining the relative degree of influence of enrollees' social location, health care needs, and health care experience, on their satisfaction with their health plan.

Figure 6 – HMO Enrollee Satisfaction: Analytical Framework



“Social Location” reflects age, gender, education, income, race, and ethnicity. All of these variables were included in the Medicaid analysis. Race and ethnicity were excluded from the commercial analysis due to the very low numbers of non-white and Hispanic enrollees in the sample.

“Health Care Need” was conceptualized in five dimensions: perceived health, physical health, mental health, and incidence of serious (physician-diagnosed) and other chronic conditions. Perceived health was based on enrollee self-rating, using a scale of poor to excellent health. Physical health and mental health were evaluated using summary scales derived from the Medical Outcomes Study Short-Form set of twelve health-related functioning (SF-12) using weights provided by the Health Institute. Information about the incidence of serious and other chronic conditions was taken from the Utah satisfaction survey.

DRAFT

“Health Care Experience” was measured in terms of health care utilization and perceived problems with access to care. Utilization was based on the number of clinic visits and hospitalization during the twelve months preceding the survey. Perceived problems with access to care were based on enrollees’ perceptions of their experience obtaining care that their physician recommended, delay in care while waiting for approval, and difficulty obtaining a referral to a specialist.

“Enrollee Satisfaction” was viewed in terms of the overall satisfaction with the health plan and as satisfaction with five domains: quality of care, coverage, appointments, provider choice, and physical access. These five domains were derived from enrollee evaluations of 22 specific aspects of the health plan or the health care they had received

Logistic regression was performed on the satisfaction variables separately, with social location, health care needs, and health care experience as independent variables. This portion of the analysis examined whether health care experience exerts a significant influence on satisfaction and if the effect remains with the inclusion of social location and health care needs variables. Two sets of analysis were performed, one for Medicaid enrollees, and one for commercial enrollees.

Tables 6 and 7 summarize the results of the analysis. The cells in the main body of the tables contain relative odds, or odds ratios. The relative odds pertain to the odds that the subgroup of beneficiaries coded as one (indicated in parentheses for each variable) reported being satisfied, relative to those coded zero (the reference subgroup). The indicated subgroup of beneficiaries is interpreted to have lesser, similar, or greater satisfaction, compared to the reference category, according to whether the odds are significantly less than one, close to one, or significantly greater than one respectively.

The lower panel of each table contains five rows of summary statistics. The first row is the model chi-square, which indicates whether the logistic regression model is statistically a

DRAFT

better fit for the data than a model that only contains a constant term. The second row contains the $2 \times \log$ -likelihood ratios which reflect the goodness of fit of the logistic regression, where smaller values mean better fit. The difference in $2 \times \log$ -likelihood ratios between two models is distributed as a chi-square with degrees of freedom equal to the difference in the degrees of freedom of the two models. Statistical significance of this difference means that the model with the smaller log-likelihood ratio is a better fit than the other. The third row of the panel contains differences and their statistical significance. The model number is indicated in parentheses. The fourth and fifth rows of the summary statistics panel show the degrees of freedom for each model and the number of valid cases.

Focusing on the panels for health care experience, having or not having an access-related problem does have a significant and independent effect on satisfaction for either Medicaid or commercial plans. This finding is consistent for overall satisfaction and for the five different domains. The results for utilization are not as consistent, but generally reflect the same finding.

In summary, the research examined the relative importance of fundamental causes (e.g., social location) and proximate determinants (e.g., health care experience) in accounting for variation in health plan satisfaction. It was proposed that both are important. The research found that while social location and health care needs exert independent effect on satisfaction with some aspects of the health plan, health care experience emerged as having strong and consistent influence on satisfaction. Health care experience was a significant contributor to the variation in satisfaction among beneficiaries, overall and domain-specific and, for the most part, within each of the four age groups. After adjusting for all other factors in the study, enrollees having experienced a problem with access to care had a decreased likelihood of being satisfied with any aspect of their plan.

DRAFT

TABLE 6

Odds Ratios from Logistic Regression of Enrollee Satisfaction with Health Plan Overall and with Domain-Specific Satisfaction - Medicaid

	Overall Satisfaction	Quality of Care	Coverage	Appointments	Provider Choice	Physical Access
Age						
18-24	0.89	0.92	0.58***	0.94	1.21	1.38
25-34	0.99	1.25	0.93	1.33*	0.75*	0.86
35-54	0.79	0.80	1.15	1.09	0.84	1.00
55 + (reference)	---	---	---	---	---	---
Social Location						
Gender (Female=1)	1.21	0.93	1.38	1.50	0.51**	1.34
Education (H.S Graduate=1)	0.90	1.16	0.93	0.98	0.59**	1.42*
Income (\$15,000+=1)	1.52	1.01	1.19	0.90	1.07	0.80
Race (White=1)	0.83	1.10	0.88	1.56	1.07	0.78
Ethnicity (Non-Hispanic=1)	1.13	2.77*	1.06	0.90	1.69	1.36
Health Care Needs						
Perceived Health (VG/E=1)	2.27**	1.74*	1.59*	1.13	1.16	1.82**
Physical Health Index (7-66)	1.01	0.97**	1.01	1.00	1.01	0.99
Mental Health Index (11-66)	1.01	1.01	1.03***	1.02*	1.01	1.00
Serious Chronic Cond. (None=1)	0.78	0.87	0.84	1.12	1.51*	0.66*
Current Conditions (<2=1)	1.96*	0.95	1.04	1.17	0.86	0.86
Health Care Experience						
HC Utilization	0.89	0.52***	0.65*	1.11	1.00	0.96
Access Problem (None=1)	4.00***	2.17***	2.38***	1.88***	1.92***	1.53*
Model Chi-square (all significantly different from Constant model)	116.9	77.7	80.5	46.8	65.3	41.0
2*Log-Likelihood Ratio	579	898	894	927	910	935
ΔL^2 relative to model without Health Care Experience	58**	50***	49***	37***	35***	28***
Degrees of Freedom	15	9	15	15	15	15
Number of Cases	614	614	614	614	614	614

=p<.05; **p<.01; ***p<.001. Only those variables for which significant effects were observed in preliminary analyses are included in the final model. Results of interaction terms included in the model, which varied by domain, are not shown. Cells in main body of the table contain odds ratios.

DRAFT

TABLE 7
Odds Ratios from Logistic Regression of Enrollee Satisfaction with Health Plan Overall
and with Domain-Specific Satisfaction - Commercial

Age	Overall Satisfaction	Quality of Care	Coverage	Provider Choice	Physical Access
18-24	1.49*				
25-34	0.92				
35-54	0.60***			3.07***	
55 + (reference)	---				
Social Location					
Gender (Female=1)				3.21***	1.93*
Education (College Graduate=1)					
Health Care Needs					
Perceived Health (VG/E=1)	2.09***	3.35***			2.73**
Physical Health Index (7-66)					
Mental Health Index (11-66)			1.22*		
Serious Chronic Condition (None=1)	1.53*	23.40*			4.17**
Current Conditions (<2=1)					
Health Care Experience					
HC Utilization	1.74*	6.90**			
Access Problem (None=1)	4.17***	16.70**	3.13**	2.56***	.24***
<hr/>					
Model Chi-square (all significantly different from Constant model)	118.1	44.4	66.7	69.4	55.2
-2*Log-Likelihood Ratio	1009	273	1220	467	272
Degrees of Freedom	8	9	6	10	7
Number of Cases	1006	793	466	964	250

=p<.05; **p<.01; ***p<.001. Only those variables for which significant effects were observed in preliminary analyses are included in the final model. Results of interaction terms included in the model, which varied by domain, are not shown. Cells in main body of the table contain odds ratios.

DRAFT

5. Conclusions

The need for reliable, instructive, measures of health care performance is increasing and new sectors of the health care industry are demanding such information. As HEDIS measures become the standard tool for health plan comparisons, this systematic evaluation of its effectiveness in discriminating between health plan performances is critical. Continuous evaluation of the accuracy of the data and a systematic assessment of their reliability and validity are essential.

In this analysis, high internal consistency was observed for most of the measure sets. This finding implies that there are opportunities for variable reduction and simplification. Parsimony in analysis could be achieved without loss of substance by recognizing measure sets that were highly internally consistent. Thus, they may be represented either by a few of the measures or a composite of several measures, depending on the relationships observed.

Taken further, the findings imply that measures within highly internally consistent measure sets may reduce the burden on resources, as not all measures need to be collected. Finally, a year-to-year rotation of measures may be instituted based on the findings. Future research that may prove worthwhile is a reconciliation of the findings of this study with the rotation scheduled instituted by the NCQA.

Significant correlations were observed in pairings between measures from one measure set with measures in another set. However, many such pairings showed no correlation, including some that were expected to be strongly correlated. An example is the absence of correlation between cholesterol screening and heart-related DRG utilization.

In general, “process” measures were found to be a weak predictor of “outcome” measures. While these findings may be an artifact of methodology, which deserves investigation,

DRAFT

they suggest the need for further examination of how process measures translate to outcomes at the health plan level.

The importance of subjective measures was illustrated by the findings that some of the measures remain correlated with subjective measures, even when adjusting for health plan type and enrollment composition. Enrollees' satisfaction with their health plan is influenced by actual experience, independent of socio-demographic characteristics and health status.

NCQA has established and implemented systematic and standardized audits of HEDIS data. Continuing these audits is important. As measures become more clinical, such an audit is essential to ensure validity of the data. For usability and meaningfulness of data, reduced measure sets and composite scores derived from the HEDIS sets are recommended.

The question of patient level data versus health plan level data should be addressed. One tool may not fit all situations. A model of reporting mix measures and different units of analysis to increase the sensitivity of the measures is needed. An integrated data set or repository, which would allow derivation of different levels of measures (enrollee level, health plan level, and provider level), is also needed.

The enormous potential of HEDIS as a tool for evaluating quality of care provided to members goes beyond its use of health plan report cards to stimulate competition and guide consumer decision-making. National databases containing years of data from many health plans covering all geographic areas in the country provide ample opportunity for analysis of:

- the factors that drive health plan performance in various domains;
- the factors that drive consumer satisfaction; and
- the relationships between health plan performance and consumer satisfaction.

This study proposes and demonstrates the application of a framework for such analysis.

DRAFT

REFERENCES

Carmines, E.G. and Zeller R.A. *Reliability and Validity Assessment*, Sage Series: Quantitative Applications in the Social Sciences, Sage Publications, Beverly Hills, CA: 1979.

Daley, "Risk Adjustment for Measuring Health Care Outcomes." edited by Liza Iezzoni, 1994.

DeVellis, R.F. *Scale Development Theory and Applications*, *Applied Social Research Methods Series*, Volume 26, Sage Publications, 1991.

Epstein, A. "Rolling Down the Runway. The Challenges Ahead for Quality Report Cards." *JAMA*, June 3, 1998, 279(21), 1691-1696.

Faulkner and Gray. *1999 Comparative Performance Data Sourcebook*. New York, NY: F & G Inc.

Faulkner and Gray. *1998 Comparative Performance Data Sourcebook*. New York, NY: F & G Inc.

Faulkner and Gray. *1997 Comparative Performance Data Sourcebook*. New York, NY: F & G Inc.

Litwin, M.S. *How to Measure Survey Reliability and Validity*. Sage Publications, Beverly Hills, CA, 1995.

Markowitz, D. "Utah and Maryland Lead with New Data Reporting Initiatives." *1998 Comparative Performance Data Sourcebook*, Faulkner & Gray, NY, NY, 1998.

DRAFT

McGlynn, Elizabeth A. and Asch, Steven M. "Developing a Clinical Performance Measure". *American Journal of Preventive Medicine*, 14 (3S), 1998.

National Committee for Quality Assurance *HEDIS 3.0*, Vol.1, NCQA, Washington DC, 1997.

Paita, Luis MC. "Utah's 1996 Survey of Enrollee Satisfaction in Medicaid-Contracted HMOs in Utah." *Utah's Health Journal*, 1997.

Vogt, Paul W. *Dictionary of Statistics and Methodology*, Sage Publications, 1993.

DRAFT

APPENDIX A

EVALUATION OF INTERNAL CONSISTENCY

The internal consistency of each measure set defined for this study was evaluated using three related and complementary methods of evaluating internal consistency.*

1. Examination of correlation (Spearman's Rho) matrices for each measure set. This provided pair-wise comparisons of the measures in the set.
2. Reliability analysis on each set, excluding measures that statistically do not belong in the set based on findings from correlation matrix.
3. Factor analysis on the measures within the measure sets to verify the results of the preceding analyses and to examine statistical clustering within sets.

A smaller, "reduced," set of measures was defined based on the second and third steps. Reduced measure sets are created as a sum or average of original measures, or through a factor analysis. The reduced measure sets are used for the evaluation of validity.

1. Quality of Care Measure Sets

A. Immunization Set

The immunization measure set includes overall immunization rate and rates of immunization for DPT, OPV, measles, mumps, rubella, and H influenza type B. While the overall immunization rate was not correlated with type-specific immunization rates, all the other immunization rates were significantly and highly correlated. This indicates high internal consistency, with an Alpha value of 0.95. The measles immunization rate correlated most highly

DRAFT

with the other immunization rates. Two measure sets, overall immunization and the average of the type-specific immunization rates, represent the immunization set in subsequent analyses.

B. Preventive Screening Set

The preventive screening set includes cholesterol, mammography, and cervical cancer screening rates. The three measures are significantly and highly correlated with each other (Alpha=0.69). The average of these three rates provided a single measure for preventive screening.

C. Asthma Set

Fifteen pair-wise comparisons of the six asthma related measures yielded significant positive correlation. Eight comparisons had a correlation of greater than .50. The computed Cronbach's alpha value (0.60) indicates high internal consistency in this set, which could justify selecting or creating a measure representative of this set. However, by adding the rates for the age groups 2-19 and 20-39 for admission rates, readmission rates, and asthmatics with more than one discharge, three rates were created. This allows use of the appropriate measure, for instance when doing analysis of all admission rates. Further reduction will be addressed later in the analysis.

* A measure set is a group of HEDIS measures that pertain to a particular area of health plan performance. Most measure sets mirror HEDIS categories and subcategories of measures. However, additional measure sets were created using hypothesized linkages between measures that cross tables

DRAFT

2. High Occurrence/High Cost DRGs, Discharges per 1,000 Measure Set

A. Discharges per 1,000 Measure Set

Eighteen of the 36 bivariate correlations among the nine DRGs in this measure set were significant. Of these 18 only two, the correlation between cerebrovascular disorders and transient ischemic attack and the correlation between simple pneumonia and pleurisy, had magnitudes greater than 0.50. This indicates low internal consistency and suggests that there is no propensity for health plans to have consistently low or high rates of discharge for these nine DRGs. That is, in general, health plans in this study do not appear to selectively attract enrollees with high (or low) probabilities of having discharges for these nine DRGs. The overall reliability of the DRG measure set is 0.56.

B. Average Cost Per Discharge Measure Set

Twenty-three of the 36 bivariate correlations among the nine DRGs in this measure set were significant. Of the 23, only three had magnitudes greater than 0.50. This indicates low internal consistency within this measure set ($\text{Alpha}=0.60$) and no propensity for health plans to have consistently low cost discharges or consistently high cost discharges for these nine DRGs. That is, in general, health plans in the data set do not appear to selectively attract enrollees based on expected discharge cost for these DRGs.

C. Average Length of Stay Per Discharge Measure Set

Twenty-seven of the 36 bivariate correlations conducted among the nine DRGs in this set were significant. Of the 27, only one was greater than 0.50. This indicates low internal consistency ($\text{Alpha}=0.51$) and no propensity for health plans to have consistently short stays or consistently long stays for these DRGs. That is, in general, health plans do not appear to selectively attract enrollees based on anticipated length of stay for these DRGs.

DRAFT

D. Within a DRG

All of the pair-wise DRG-specific correlations between cost per discharge and average length of stay were significant. Six such correlations were greater than 0.50. Of the nine DRG-specific correlations between discharge rate and average length of stay, eight were significant and three were greater than 0.50.

There were no consistently high correlations between discharge rate and average cost per discharge across DRGs or by specific DRG. Of the nine DRG-specific correlations, only four were significant and three were greater than 0.50.

E. Selected Procedures Set (per 1,000 members)

The selected procedures set includes: (1) sex-specific angioplasty, cardiac catheterization and CABG rates for enrollees aged 45-64, cholecystectomy for enrollees age 30-64, and laminectomy/discectomy for enrollees aged 20-64; (2) hysterectomy rates for females aged 15-44 and 45-64; and (3) prostatectomy rates for males 45-64.

Internal consistency within the whole set is high. All of the pair-wise correlation coefficients were statistically significant. Particularly high internal consistency was observed among heart-related measures that were linearly dependent (based on linear regression).

The sex-specific measures showed high correlation between male and female rates for each procedure. An average of the male and female rates for each procedure was computed for set reduction.

3. Inpatient Utilization - General Hospital/Acute Care Measure Set

This measure set includes average length of stay calculated separately for each age group 0-19, 20-44, 45-64, 65+ and for all ages combined. Discharge rates are for all types of

DRAFT

discharges—medical/surgical and maternity. The focus in evaluating this set was to examine the correlation of measures between age groups within utilization category, and between utilization categories within each age grouping.

A. Total Utilization

There was high internal consistency in this set ($\text{Alpha}=0.72$). All pair-wise correlations for age-specific measures and overall measures are statistically significant. The overall measure was strongly correlated with each age-specific measure. A linear regression between the overall measure and the age-specific measures indicated that the age-specific measures were significantly linearly related to the overall measure, except for age group 65+. In the multivariate analyses, the overall measure represented the under age 65 age specific groups measure set. Age group 65+ measures were treated separately in further analyses.

A. Medical/Surgical

As in the results for total discharges, all pair-wise correlations between age-specific measures and the overall measure were statistically significant. In the multivariate analyses, the overall measure is used to represent this set. Again, the measure for age group 65+ is treated separately in further analyses.

B. Maternity

The results for this set differ slightly from those for medical/surgical discharges. The overall measure is strongly correlated with the measures for age groups 0-19 and 20-44, but not with the measures for age groups 45-64 and 65+ (although the correlation with age group 45-64 is statistically significant). Again, in the multivariate analyses, the overall measure is used to represent this set with the measure for age group 65+ treated separately in further analyses.

C. Age 0-19 Average Length of Stay (ALOS)

DRAFT

Statistically significant and high correlation was observed between the overall ALOS and the medical/surgical ALOS within age group 0-19. The correlation between the overall ALOS and the maternity ALOS was statistically significant but not high. No statistically significant correlation was observed between the medical/surgical ALOS and the maternity ALOS.

D. Age 20-44 Average Length of Stay (ALOS)

Statistically significant and high correlations were observed between the overall ALOS and the medical/surgical ALOS and between the overall ALOS and the maternity ALOS within age group 20-44. The correlation between the medical/surgical ALOS and the maternity ALOS was statistically significant but not high.

E. Age 45-64 Average Length of Stay (ALOS)

Statistically significant and high correlation was observed between the overall ALOS and the medical/surgical ALOS within age group 45-64. The correlation between the overall ALOS and the maternity ALOS, and between the medical/surgical ALOS and the maternity ALOS, were also statistically significant but not high.

F. Age 65+ Average Length of Stay (ALOS)

Statistically significant and high correlation was observed between the overall ALOS and the medical/surgical ALOS within age group 65+. No statistically significant correlation was observed between the overall ALOS and the maternity ALOS. The correlation between the medical/surgical ALOS and the maternity ALOS was statistically significant but not high.

DRAFT

G. All Ages Average Length of Stay (ALOS)

All three pair-wise correlation coefficients for the ALOS measures for all ages were statistically significant and high, indicating that the ALOS for all discharges for all ages would adequately represent this set. Newborn discharge rate and newborn days are significantly statistically correlated.

4. Ambulatory Care Measure Set

The ambulatory care measure set includes measures calculated separately for each age group 0-19, 20-44, 45-64, and 65+, as well as for all ages combined. Three groups of utilization measures (on a per 1,000 member basis) are in this set; outpatient visits, emergency room visits, and ambulatory surgery procedures. Our focus in evaluating this set is to examine the correlation of measures between age groups within utilization category.

A. Outpatient Visits

All age-specific measures are statistically and highly correlated. Reliability analysis shows that higher Alpha value is achieved with the exclusion of the measure for age group 65+ (an increase from 0.79 to 0.93). The overall measure will be adequate to represent this set. However, results indicate that it might be informative to treat the measure for age group 65+ separately.

B. Emergency Room Visits

All pair-wise correlations for the age-specific measures are statistically significant. However, correlations between the measure for age group 65+ and the rest of the measures are substantially lower than the other pair-wise correlations. Reliability analysis shows that a higher Alpha value is achieved with the exclusion of the measure for age group 65+ (an increase from

DRAFT

0.86 to 0.93). The overall measure will be adequate to represent this set. However, results indicate that it might be informative to treat the measure for age group 65+ separately.

C. Ambulatory Surgery

Most of the pair-wise correlations for the age-specific measures are statistically significant. However, correlations between the measure for age group 65+ and the rest of the measures are substantially lower than the other pair-wise correlations. Reliability analysis shows that a higher Alpha value is achieved with the exclusion of the measure for age group 65+ (an increase from 0.70 to 0.89). The overall measure will be adequate to represent this set. However, results indicate that it might be informative to treat the measure for age group 65+ separately.

5. Inpatient Utilization Non-Acute Care

The non-acute, inpatient utilization measure set includes measures calculated for each age group 0-19, 20-44, 45-64, and 65+, as well as for all ages combined. Three groups of utilization measures are in this set: discharges per 1,000 members, days per 1,000 members, and average length of stay. Our focus in evaluating this set is to examine the correlation between age groups for each utilization measure, and between utilization measures within each age grouping.

A. Discharges per 1,000 Members

All pair-wise correlation coefficients were significant. Specifically, high correlations were observed between measures for adjacent age groups. The measure for all-ages was highly correlated with all age-specific measures and can represent the set. However, reliability analysis reveals that the Alpha value is low (0.37) with the measure for age 65+ included and substantially high (0.63) when it is excluded.

DRAFT

B. Days per 1,000 Members

All pair-wise correlation coefficients were significant and greater than 0.50. The all-ages measure was highly correlated with all age-specific measures and can represent the set. However, as with discharges per 1,000 members, reliability analysis reveals that the Alpha value is low (0.23) when the age 65+ measure is included and substantially high (0.65) when it is excluded.

C. Average Length of Stay

All pair-wise correlation coefficients were significant. Unlike discharges and days per 1,000, the all-ages measure was highly correlated only with measures for ages 20-44 and 45-64. Reliability analysis reveals that the Alpha value is high (0.80) with all measures included in the set.

6. Utilization Measures Within Age Groups

For each age group, and for all ages combined, correlation and reliability analyses were performed for outpatient visits, emergency room visits, ambulatory surgery procedures, non-acute inpatient discharge rates and inpatient non-acute average length of stay.

A. Age 0-19

Age 0-19 indicated low internal consistency. Of the 10 pair-wise correlation coefficients examined, only three were significant. Of these, the only high correlation was between inpatient non-acute discharge rates and inpatient non-acute average length of stay.

DRAFT

B. Age 20-44

Of ten pair-wise correlation coefficients, only five were statistically significant. Of these, the only high correlation was between inpatient non-acute discharge rates and inpatient non-acute average length of stay. There was modest correlation between emergency room visits and ambulatory surgery procedures per 1,000.

C. Age 45-64

The results for this age group were similar to those for age group 20-44: low internal consistency. Although seven of the ten pair-wise correlation coefficients were statistically significant, the only high correlation was between inpatient non-acute discharge rates and days per 1,000 members. As with the age group 20-44, there was modest correlation between emergency room visits and ambulatory surgery procedures per 1,000 members.

D. Age 65+

A high internal consistency was observed for this group. All ten pair-wise correlation coefficients were statistically significant and most were greater than 0.60. The greatest correlation was between inpatient non-acute discharge rates and days per 1,000 members.

E. All Ages

Across all ages there was low internal consistency. Of ten pair-wise correlation coefficients, six were statistically significant. Again, the only high correlation was between inpatient non-acute discharge rates and days per 1,000 members. There was modest correlation between emergency room visits and ambulatory surgery procedures per 1,000 members.

DRAFT

7. Discharges and Average Length of Stay for Females in Maternity Care Measure Set

This set includes discharges per 1,000 female members, days per 1,000 female members, and average length of stay for age groups 10-14, 15-34, and 35-49, as well as for all ages. Measures were also calculated for age groups 10-49 and 15-49. Maternity care measures are in four categories: total deliveries, vaginal deliveries, cesarean section, and vaginal delivery after cesarean section.

The objective in evaluating this set was to examine internal consistency between age-specific measures within each type of delivery, and internal consistency between types of delivery within each age group.

A. Total Deliveries per 1,000 Members

Four of the six pair-wise correlation coefficients from the all-ages and four age-specific measures were significant. The all-ages measure was highly correlated with measures for ages 15-34 and 35-49 but not with 10-14. The Alpha value is 0.65 for all-ages, 15-34, and 35-49 measures.

The all-ages measure may adequately represent the measure for women of childbearing ages (15-49).

B. Vaginal Deliveries (Per 1,000 Live Births)

Three of the six pair-wise correlation coefficients from the all-ages, and three age-specific measures, measures for ages 10-14, 15-34 and 35-49, were significant. The three non-significant correlations were between the measure for age 10-14 and the other age groups. The all-ages measure was highly correlated with measures for ages 15-34 and 35-49 but not with 10-14. The Alpha value was 0.65 for all-ages, 15-34, and 35-49 measures.

DRAFT

The all-ages measure may adequately represent the measure for women of childbearing ages (15-49).

C. Cesarean Section

Three of the six pair-wise correlation coefficients from the all-ages and three age-specific (10-14, 15-34, 35-49) measures were significant. The three non-significant correlations were between the measure for age 10-14 and the other age groups. The all-ages measure was highly correlated with measures for ages 15-34 and 35-49 but not with 10-14. The Alpha value is 0.80 for all-ages, 15-34, and 35-49 measures.

The all-ages measure may adequately represent the measure for women of childbearing ages (15-49).

8. Newborn Measure Set

The newborn measure set includes volume (live births per 1,000 female members) and average length of stay for all newborns, well newborns, and complex newborns per 1,000 female members.

There was high correlation between volume and length of stay measures for all newborns with the respective measures for well newborns and complex newborns. However, the correlation between the measures for well newborns and complex newborns, while statistically significant, is low. The measures for all newborns can adequately represent this set for subsequent analyses.

9. Mental Health Utilization Measure Set

The mental health measure set includes separate mental health utilization measures for male and female members. The objective in evaluating internal consistency in this set was to

DRAFT

identify the correlation between the male and female measures for each mental health service category and the correlation between service types and gender, for males, females, and all members.

A. Inpatient Discharges Per 1,000 Members

There was high internal consistency among male, female, and both-sexes measures. The measure for all members adequately represents this subset.

B. Inpatient Average Length of Stay

There was high internal consistency among male, female, and both-sexes measures. The measure for all members adequately represents this subset.

C. Percent Of Members That Receive Mental Health Service

There was high internal consistency among male, female, and total member measures. The measure for all members adequately represents this subset.

D. Percent Of Members Receiving Inpatient Mental Health Services

There was high internal consistency among male, female, and both-sexes measures. The measure for all members adequately represents this subset.

E. Percent Of Members Receiving Day/Night Mental Health Services

There was high internal consistency among male, female, and both-sexes measures. The measure for all members adequately represents this subset.

DRAFT

F. Percent Of Members Receiving Ambulatory Mental Health Services

There was high internal consistency among male, female, and both-sexes measures. The measure for all members adequately represents this subset.

H. Mental Health Measures Evaluated as a Set

The internal consistency of this set is high (Alpha = 0.80) when all mental health measures, discharge rates, ALOS, percent receiving various mental health services, and readmission rates, are included in the analysis as a set.

10. Readmission for Major Affective Disorder Measure Set

A. Percent Of Members Readmitted Within 90 Days

A statistically significant but weak correlation (Rho= 0.42) was observed between male and female measures. The all member measure was highly correlated with both the male-specific and the female-specific measures.

B. Percent Of Members Readmitted Within 365 Days

A statistically significant but weak correlation (Rho= .42) was observed between measures for males and females. However, the all members measure was highly correlated with both the male-specific and the female-specific measures.

C. Readmission Within 90 Days

Pair-wise correlations between 90-day readmission rates and 365-day readmission rates for men, women and all members were significant. However, low correlation was observed for the female – only measures.

DRAFT

11. Chemical Dependency - Utilization Measures Measure Set

The chemical dependency measure set includes separate chemical dependency utilization measures for male and female members. The objective in evaluating the internal consistency of this set was to identify a correlation between the male and female measures for each service category, and a correlation among service types for males, for females, and for all members, respectively.

A. Inpatient Discharges Per 1,000 Members

There was high internal consistency among male, female, and both-sexes measures with respect to inpatient discharges per 1,000 members. The measure for all members adequately represents this subset.

B. Inpatient Average Length Of Stay

There was high internal consistency among male, female, and both-sexes measures. . The measure for all members adequately represents this subset.

C. Percent Of Members Receiving Any Chemical Dependency Services

There was high internal consistency among male, female, and both-sexes measures. . The measure for all members adequately represents this subset.

D. Percent Of Members Receiving Inpatient Chemical Dependency Services

There was high internal consistency among male, female, and both-sexes measures. . The measure for all members adequately represents this subset.

DRAFT

E. Percent Of Members Receiving Day/Night Chemical Dependency Services

There was high internal consistency among male, female and all member measures. The measure for all members adequately represents this subset.

F. Percent Of Members Receiving Ambulatory Chemical Dependency Service

There was high internal consistency among male, female and all member measures. The measure for all members adequately represents this subset.

G. Chemical Dependency Measures Evaluated as a Set

The internal consistency was high ($\text{Alpha}=0.90$) when all chemical dependency measures--discharge rates, ALOS, percent receiving various chemical dependency services, and readmission rates--were included as a set in the analysis.

12. Readmission for Chemical Dependency Measure Set

A. Percent Of Members Readmitted Within 90 Days

There were statistically significant and high correlation among male-specific and female-specific measures and the measure for both sexes.

B. Percent Of Members Readmitted Within 365 Days

There was a statistically significant but weak correlation ($\text{Rho}=0.31$) between male and female measures. The all member measure was highly correlated with both the male and the female specific measures.

DRAFT

C. Readmission Within 90 Days versus Readmission Within 365 Days

All three (male, female, both sexes) pair-wise correlations between 90-day and 365-day readmission rates were statistically significant. However, the correlation coefficient was less than .51 in all three cases.

13. Outpatient Drug Utilization Measure Set

Average cost of prescriptions on a per member per month basis (PMPM) and average number of prescriptions (PMPM) were highly correlated (coefficient close to 1.0).

DRAFT

APPENDIX B

1996 UTAH SURVEY OF ENROLLEES IN MEDICAID-CONTRACTED HMOS

I. Survey Planning And Implementation

The 1996 Utah Survey of Enrollees in Medicaid-Contracted HMOs is part of Utah's health care performance measurement project. The project uses "report cards," publicly disseminated measures of performance by health plans and providers of health care, to create competition based on quality of care. The publication of these health plan performance measures is expected to create incentives for improvement in quality of health care.

The Medicaid enrollee survey had three objectives: (1) to collect information that allows objective comparison of health plans according to various satisfaction measures, (2) to identify the features of care and service that contribute most to enrollee satisfaction, and (3) to examine how evaluations of satisfaction differ among subgroups of enrollees. Subgroups are defined, for example, by demographic characteristics, socioeconomic status, health status, or utilization patterns.

The Survey of Enrollees in Medicaid-Contracted HMOs was conceived, planned, and implemented with the voluntary participation of the five Medicaid-contracted HMOs - United HealthCare, Intermountain Health Care of Utah, Family Health Plan of Utah, Intergroup Healthcare, and Utah Blue Cross/Blue Shield*. The five HMOs enroll 88 % of Utah's HMO market (663,360 enrollees in July, 1996), 82 % of Utah's managed care market, and 100 % of Utah's Medicaid, managed care market*.

* In 1996, Intergroup of Utah ceased to be a contractor for the Medicaid program

*Based on data from Utah Association of Health Care Providers, Eye on the Market, 1996.

DRAFT

The Utah Department of Health's Office of Health Care Statistics (formerly known as the Office of Health Data Analysis) coordinated survey planning.

J. Survey Instrument

The survey instrument was developed through consultation with representatives from HMOs, Division of Health Care Financing, Office of Health Data Analysis, Health Data Committee, and other Department of Health staff. Of the five HMOs, three already used the NCQA instrument, one used the Group Health Association of America instrument (now known as the American Association of Health Plans), and one used none of the nationally recognized instruments. Agreement was reached to use the NCQA instrument as the basis for the survey instrument with some modification.

The survey instrument contained the following items relating to the performance and quality of the health plan:

1. Satisfaction ratings, specific aspects of care/plan:

K. Quality of care, including:

- a. Thoroughness of treatment
- b. Attention to what an enrollee has to say
- c. Amount of time with doctors or staff
- d. Outcomes of an enrollee's medical care
- e. How well an enrollee's needs are met
- f. How well different people and departments communicate
- g. Overall quality of care
- h. Thoroughness of exam and accuracy of diagnosis
- i. Thoroughness of explanations

DRAFT

- j. Friendliness of doctors and staff
- k. Advice about ways to avoid illness
- l. Sensitivity to cultural or religious background

L. Plan coverage and information, including:

- a. Range of services covered by health plan
- b. Information about covered services
- c. Coverage for preventive care
- d. Availability of medical advice by telephone

M. Appointments:

- a. Ease of making an appointment
- b. Time between making the appointment and the appointment itself

N. Provider choice:

- a. Number of doctors
- b. Ease of choosing a personal physician

O. Physical access:

- a. Convenience of the location of the doctor's office
- b. Evening and weekend access

2. **Overall measures of satisfaction and perceived health plan quality and performance:**

- Overall satisfaction
- Intention to switch
- Would recommend to family or friends
- Perceived change in overall performance

DRAFT

3. Perceived problems:

- Delays in getting care while waiting for approval
- Not getting care that doctor believes is necessary
- Difficulty in getting referral to specialist desired

4. Other experiences related to health plan performance:

- Making appointments
- Waiting time between appointment and actual visit
- Waiting time in the provider's office
- Lodging complaints
- Resolution of complaints

The survey also collected information on enrollees' health status, socio-demographic characteristics, health care utilization and length of enrollment with the health plan.

P. Survey Sample

The survey sample consisted of five strata, the five Medicaid contracted HMOs. Systematic sampling with random start was employed to generate a representative sample from each stratum.

The sampling design apportioned the interviews of 400 Medicaid enrollees per plan for a total of 2,000 Medicaid enrollees overall. This sample allocation provided enough analytic power to draw conclusions about each health plan's Medicaid enrollees. The following enrollees were selected: Utah residents, currently enrolled in the health plan as of May 1, 1996, with no member of the family employed by the health plan, and with phone numbers on record. Additional selection criteria were:

DRAFT

- Not in foster care, group homes, or youth correctional facility;
- Continuously enrolled for at least three months; and
- Not enrolled in Medicare or any commercial health plan.

Q. Survey Implementation

The survey was conducted by telephone from DataStat's central location Computer Assisted Telephone Interviewing (CATI) facility in Ann Arbor, Michigan. DataStat is an NCQA-certified survey agency selected by Office of Health Care Statistics through a competitive bid process. Interviews were conducted during the evenings and/or on weekends.

The number of enrollees located per 100 phone numbers called ranged from 65.6 to 73.0 across the five strata. The average contact rate was 68.7 percent. Response rates, the number of enrollees who completed the interview per 100 enrollees determined to be eligible for the survey, ranged between 76.8 and 83.2 percent or 79.5 percent overall. Among Medicaid populations, the contact rate is usually low due to the mobility of the population and the low rate of household phones. Once contacted, most enrollees are willing to participate.

The enrollees who completed the interview were compared to the enrollees in the sampling pool in terms of sex, age, and duration of enrollment (the only available information). There was some evidence of selectivity by age: enrollees who completed the interviews were younger than those in the sampling pool. Based on this finding, an adjustment factor for the enrollee population's age distribution was calculated and applied to the HMO sample.

Because of the age-based criteria used for the selection of a respondent and deciding whether to use a proxy, the enrollees who responded to the interviews themselves differ significantly from those for whom a proxy respondent completed the interview in terms of socio-demographic characteristics and health status. (Preliminary analysis is not shown here.) Because of these differences, there is a possibility of bias introduced by the use of proxy

DRAFT

respondents. This is not a concern for this specific study, which focuses on enrollees aged 18 years and older and enrollees who themselves completed the interviews.

R. Weights and Adjustment Factors

A sampling weight was calculated for each of the five strata to compensate for the disproportionate sampling (equal-size across HMOs). The weights were calculated based on the Medicaid enrollment as of May 1996. The weighted distribution of the sample by HMO reflects their relative market share in the Medicaid-contracted HMO market in Utah. Based on the results of the test for sample selectivity, an adjustment factor was calculated to adjust the samples to reflect the age distribution of the sampling pool of eligible enrollees as submitted by the HMOs to DataStat. There was no evidence of sensitivity to the weighting scheme applied (unweighted, weighted for disproportionate sampling, and weighted to reflect age distribution) of responses to the satisfaction items or health status. Age-adjustment factors were calculated to adjust the sample to the age-distribution of each HMO's enrollment population. In all stages of the analysis in the study, sampling weights and age-adjustment factors were applied.

DRAFT

APPENDIX C

Glossary

Alpha

In hypothesis testing, the Alpha level, or α -level, is the threshold for defining statistical significance. For example, setting Alpha at a level of 0.05 implies that investigators accept that there is a 5% chance of concluding incorrectly that an intervention is effective when it has no true effect.

Factor Analysis

Factor analysis is a statistical method used to reduce a large set of variables into a smaller set of factors or latent variables.

Health Plan Employer Data and Information Set (HEDIS)

Developed by the National Committee for Quality Assurance, HEDIS is a set of standardized performance measures designed to enable purchasers to evaluate a managed care plan's administrative performance. Introduced in 1993 as HEDIS 2.5, the data set has since been revised and enhanced to include more measurements that focus on the managed care plan's provision of preventive care and other proxies for quality. HEDIS also includes a standardized survey of consumers' experiences that evaluates plan performance in areas such as customer service, access to care and claims processing.

DRAFT

Internal Validity

Internal validity is the extent to which the findings of a study accurately represent the causal relationship between an intervention and an outcome in the particular circumstances of the study.

Managed Care

Managed care is a health care network in which costs are controlled through utilization management. Primary care providers serve as the “coordinators” for patients’ access to specialty care. **Managed care systems** are systems of health care delivery that influence utilization and cost of services and measure performance. **A managed health care plan** is one or more products that integrates financing and management with the delivery of health care services to an enrolled population. A managed health care plan employs or contracts with an individual provider(s) or organized provider network that delivers services and has a financial risk or incentive to deliver quality, cost-effective services. A managed health care plan uses an information system capable of monitoring and evaluating patterns of covered persons’ use and cost of services.

Types of managed care organizations include Health Maintenance Organizations (HMOs) and Preferred Provider Organizations (PPOs).

Measure Set

A measure set is a group of HEDIS measures that pertain to a particular area of health plan performance.

National Committee for Quality Assurance (NCQA)

The NCQA is an independent, non-profit organization. The NCQA works to improve

DRAFT

health care quality by providing tools to evaluate managed care plan performance and programs for the accreditation of those plans that meet specified performance standards. Participation in the NCQA's accreditation and certification programs is voluntary. For additional information about the NCQA, please go to <http://www.ncqa.org/about>

Quality Assessment

Quality assessment is a measurement and monitoring function of quality assurance for determining how well health care is delivered in comparison with applicable standards or acceptable bounds of care.

Quality Assurance

Quality assurance is activity intended to ensure that the best available knowledge concerning the use of health care to improve health outcomes is properly implemented. This involves the implementation of health care standards, including quality assessment and activities to correct, reduce variation in or otherwise improve health care practices relative to those standards.

Quality Compass

Quality Compass is a database that can be accessed by consumers, employers and other health care purchasers and used to compare HEDIS data for managed care plans across the country. Developed and sold by the National Committee for Quality Assurance (NCQA) the database includes NCQA accreditation status, HEDIS results and national averages for performance indicators.

DRAFT

Reliability

Reliability can be defined as the extent to which an observation that is repeated in the same stable population yields the same result (i.e. test-retest reliability). Also, it is the ability of a single observation to distinguish consistently among individuals in a population.

Validity

Validity is the extent to which a measure accurately reflects the concept that it is intended to measure.